

BEST AVAILABLE COPY

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 04-281559

(43)Date of publication of application : 07.10.1992

(51)Int.Cl.

G06F 15/40

(21)Application number : 03-069322

(71)Applicant : TOSHIBA CORP
TOSHIBA AVE CORP

(22)Date of filing : 08.03.1991

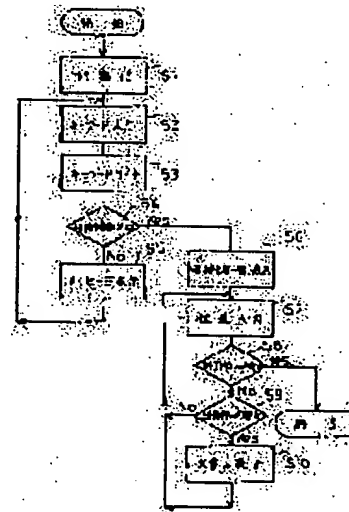
(72)Inventor : TANOSAKI YASUO
KOBAYASHI KENICHIRO

(54) DOCUMENT RETRIEVING DEVICE

(57)Abstract:

PURPOSE: To easily retrieve an objective document from a document data base.

CONSTITUTION: After changing a display configuration indicating the content of a candidate document according to the number of the candidate documents, parts including key words in text data are extracted as the element of a candidate document list, and displayed on a display screen (step S6). One of the elements of the displayed candidate document list is designated by an inputting device (step S7). The content of document data corresponding to the designated document is displayed on the display screen (step S10).



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平4-281559

(43) 公開日 平成4年(1992)10月7日

(51) Int.Cl.³

識別記号

庁内整理番号

F I

技術表示箇所

G 0 6 F 15/40

5 0 0 G 7056-5L

審査請求 未請求 請求項の数 1 (全 9 頁)

(21) 出願番号

特願平3-69322

(22) 出願日

平成3年(1991)3月8日

(71) 出願人 000003078

株式会社東芝

神奈川県川崎市幸区堀川町72番地

(71) 出願人 000221029

東芝エー・ブイ・イー株式会社

東京都港区新橋3丁目3番9号

(72) 発明者 田野崎 康雄

神奈川県川崎市幸区柳町70番地株式会社東芝柳町工場内

(72) 発明者 小林 賢一郎

東京都港区新橋3丁目3番9号東芝オーディオ・ビデオエンジニアリング株式会社

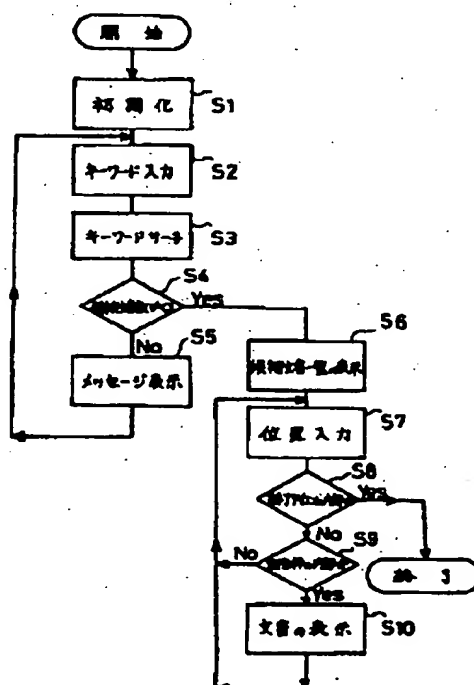
(74) 代理人 弁理士 大胡 典夫

(54) 【発明の名称】 文書検索装置

(57) 【要約】

【目的】 文書データベースの中から目的とする文書を容易に検索する。

【構成】 テキストデータ中のキーワードを含む部分を、候補文書の数に応じて候補文書の内容を示す表示形態を変化させて、候補文書一覧の要素として抽出し、表示画面に表示する。(ステップS6)。表示されている候補文書一覧の要素の中から要素の一つを入力装置により指定する。(ステップS7)。指定された文書に対応する文書データの内容を表示画面に表示する。(ステップS10)。



1

【特許請求の範囲】

【請求項1】 テキスト・データや図表データなどからなる文書データを格納する文書データ格納手段と、この文書データ格納手段に格納されている文書データを検索するキーワードを入力するキーワード入力手段と、このキーワード入力手段から入力されたキーワードを含む文書を上記文書データ格納手段の中から検索するキーワードサーチ手段とを備えた文書検索装置において、上記キーワードサーチ手段によって得られる候補文書データの数に応じて各候補文書データの内容を示す形態を変化させて該当するキーワードを含む部分を候補文書リストの要素として表示する候補文書一覧表示手段と、この候補文書一覧表示手段で表示された文書一覧における上記要素の一つを指定する文書選択手段と、この文書選択手段で指定された文書に対応する文書データの内容を表示する文書表示手段とを具備したことを特徴とする文書検索装置。

【発明の詳細な説明】

【発明の目的】

【0001】

【産業上の利用分野】 本発明は、文書データベースの中からユーザの目的とする文書を効率よく検索することが可能な文書検索装置に関する。

【0002】

【従来の技術】 大型コンピュータあるいはワークステーションを用いた文書検索システムが実用化されている。

【0003】 このような文書検索装置において文書の検索を行なう場合には、まずユーザはキーワードを入力する。その後、装置側が入力されたキーワードを、本文中に含んでいるか、あるいは検索キーとしてヘッダ部分に含んでいる文書をデータベースの中から探し出し、その検索結果をユーザに与える。

【0004】 ところで、条件を満たす文書が複数個見つかった場合には、ユーザはさらにこのなかから必要なものを選び出す必要がある。そのため、装置側は、探し出された各文書のタイトルおよび各文書に付属する文書情報あるいはアブストラクトなどの文書内容リストを文書番号とともに列挙表示し、ユーザはここに付加されている文書内容を参照して、各文書が目的にあったものか否かの判断を行なってから文書本体を閲覧している。

【0005】

【発明が解決しようとする課題】 上記したように、従来の検索装置においては、候補文書が複数ある場合に、装置側が与えた文書内容リストなどを参照して、ユーザが必要なものを選択するという形態が採られているが、文書内容リストが文書の内容を的確に表現しているケースが少なく、また、ユーザの必要とする記述が本文中に存在してもそれが文書のタイトルあるいはヘッダ情報に表されていないケースもあった。特に、候補文書数が増えた場合には、目的とする文書を検索するまでに要するユ

2

ーザの負担は大きかった。また、文書内容リスト中に詳しく各文書の内容を表現すると、文書内容リストの表示量自体が大きくなり、表示画面の表示領域に収まらず、ユーザは画面のスクロールなどを頻繁に行なわなければならないといった操作上の不具合も生じていた。

【0006】 本発明は、上記事情に鑑みてなされたもので、文書内容リスト中に各文書の内容を的確に、かつ最少限の記述量で表現できる文書検索装置を提供することを目的とする。

10 【0007】 【発明の構成】

【0008】

【課題を解決するための手段】 本発明は、上記目的を達成するために、テキスト・データや図表データなどからなる文書データを格納する文書データ格納手段と、この文書データ格納手段に格納されている文書データを検索するキーワードを入力するキーワード入力手段と、このキーワード入力手段から入力されたキーワードを含む文書を上記文書データ格納手段の中から検索するキーワードサーチ手段とを備えた文書検索装置において、上記キーワードサーチ手段によって得られる候補文書データの数に応じて各候補文書データの内容を示す形態を変化させて該当するキーワードを含む部分を候補文書リストの要素として表示する候補文書一覧表示手段と、この候補文書一覧表示手段で表示された文書一覧における上記要素の一つを指定する文書選択手段と、この文書選択手段で指定された文書に対応する文書データの内容を表示する文書表示手段とを具備したことを特徴とする。

【0009】

【作用】 本発明は上記のように構成したので、キーワードを用いることにことによって得られた複数の候補文書データの中から目的とするものを選ぶ場合に、候補文書リストの要素としてキーワードをテキスト中の周囲の語と対応づけて表示することにより、文書中でのそのキーワードの現われ方が明示表現される。

【0010】 また、候補文書リストの要素の表示形態を候補文書データの数に応じて変化させることにより、表示画面上での占有面積が小さいのかかわらず、ユーザが目的とする文書を正しく判断するに足る表示となる。

【0011】

40 【実施例】 以下、図面を参照して本発明の実施例を説明する。

【0012】 図1は、本発明の一実施例の文書検索装置の構成を示すブロック図である。

【0013】 同図に示すように、文書検索装置は、入力装置1、表示装置2、文書データ格納装置3、制御装置4、およびメモリ5から構成される。

【0014】 入力装置1は、文字コード・制御コマンド・位置情報などを入力する装置で、例えばキーボード1aとマウス1bおよびこれらを制御する装置で構成される。

50 【0015】 表示装置2は、ユーザに入力を行なわせる

3

ためのプロンプトメッセージ、入力された文字列、あるいは検索の後に得られた文書データなどの表示を行なうものであり、例えばVRAMと、このVRAMに格納されたビット情報をドット列として表示するためのディスプレイからなっている。

【0016】文書データ格納装置3は、各文書データを格納するためのものであり、例えばハードディスク装置などからなる。この文書格納装置3における文書データの格納形式を図2に示す。1個の文書データは、文書中のテキスト情報のみを含むテキストデータ部3aとイメージデータ、フォーマット情報などを含む非テキストデータ部3bからなり、文書データ格納装置3にはこのような形式の文書データが複数個格納されている。すなわち、複数の文書データ31, 32, ..., 3nは、それぞれテキストデータ部31a, 32a, ..., 3naと非テキストデータ部31b, 32b, ..., 3nbからなる形式で文書データ格納装置3に格納されている。

【0017】制御装置4は、例えばCPUなどからなるもので、入力装置1、表示装置2、文書データ格納装置3、およびメモリ5とバスにより接続されており、各装置の制御、装置間のデータの転送などの制御や処理を行なうものである。

【0018】メモリ5は、例えばダイナミックRAMからなり、図3に示すように、制御装置4が各種制御や処理を実行するためのプログラムを格納するプログラム部5aと、処理の際に必要なデータをバッファするバッファ部5bとからなっている。さらに、プログラム部5aは、メイン処理部5c、初期化部5d、キーワード入力部5e、キーワードサーチ部5f、候補文書一覧表示部5g、文書選択部5h、および文書表示部5iのモジュールに分割され、また、データバッファ部5bは、キーワード格納バッファ5j、キーワードサーチ用バッファ5k、候補文書格納バッファ5l、および候補文書数格納バッファ5mから構成される。以下、プログラム部5aとバッファ部5bの各部の機能について説明する。

【0019】メイン処理部5cは、装置全体の処理の制御を司るものであり、プログラムの分岐、初期化部5d以下の各モジュールの呼び出し（起動）などを行ない、また、初期化部5dは、各ハードウェア装置の初期設定およびデータバッファ部5bを構成する各バッファの内容の初期化を行なう。

【0020】キーワード入力部5eは、入力装置1のキーボード1aを介してユーザに検索の際にキーとなるキーワードである文字列を入力させ、これをキーワード格納バッファ5jに格納する。

【0021】キーワードサーチ部5fは、文書データ格納装置3に格納されている文書データを格納されている順序で読み出してキーワードサーチ用バッファ5kに格納し、キーワード格納バッファ5lに格納されている文字列を含む文書データをキーワードサーチ用バッファ5k上で

4

捜し出す。この検索の結果、得られる複数の文書データを候補文書データとして候補文書格納バッファ5lに格納する。

【0022】候補文書一覧表示部5gは、候補文書格納バッファ5lに格納されている各候補文書データの内容を表わす表現（以下、文書内容表現と称す）を表示装置2の表示画面上に列挙表示する。すなわち、文書内容表現は、候補文書一覧の要素として表示画面上に列挙表示される。

【0023】文書選択部5hは、すでに候補文書一覧表示部5gによって列挙表示されている文書内容表現のいずれか一つをユーザに選択させる。

【0024】文書表示部5iは、文書選択部5hによって選択された文書内容表現に対応する文書データを候補文書格納バッファ5lより読み出し、テキスト・図表などを表示装置2の表示画面上に表示する。

【0025】なお、候補文書数格納バッファ5mは、候補文書格納バッファ5lに含まれる文書データ数を格納するバッファである。

【0026】次に、上記構成の文書検索装置の具体的な処理動作について、図4に示す処理の流れのフローチャートを参照し説明する。

【0027】処理全体の制御はメイン処理部5cが司っており、メイン処理部5cはまず初期化部5dを起動する。起動された初期化部5dはバッファ部5bのキーワード格納バッファ5j、キーワードサーチ用バッファ5kおよび候補文書格納バッファ5lの初期化、候補文書数格納バッファ5mの内容のクリア、入力装置1と表示装置2の初期設定などを行なう。さらに、コマンド入力のために必要な各種のアイコンの表示も行なう。（ステップS1）。

【0028】続いて、メイン処理部5cはキーワード入力部5eを起動する。起動されたキーワード入力部5eはユーザにキーワードを入力させる。入力されたキーワードに対して、カナ漢字変換などの処理を施し、得られた文字列をキーワード格納バッファ5jに格納する。キーワードが入力されキーワード格納バッファ5jに格納された後、処理はステップS3に移行する。（ステップS2）。

【0029】ステップS3ではキーワードサーチ部5fが起動される。起動されたキーワードサーチ部5fは、文書データ格納装置3に格納されている文書データを格納されている順序、例えば最初に文書データ31を読み出し、キーワードサーチ用バッファ5kに格納する。さらに、キーワードサーチ部5fは、キーワードサーチ用バッファ5kに格納されている文書データ31のテキストデータ部31aを参照し、この中にキーワード格納バッファ5jに格納されている文字列と同一の文字列が含まれているか否かを調べる。含まれている場合には、キーワードサーチ用バッファ5kに格納されている文書データ31全体を候補文書格納バッファ5lに候補文書として格納し、候補文書数格納バッ

5

ファ5mの内容を“1”増加させる。続いて、キーワードサーチ部5fは、文書データ32から文書データ3nまでの文書データに対して上記した一連の処理を順次実行する。すなわち、文書データ格納装置3に格納されている全ての文書データに対して上記処理を実行する。(ステップS3)。

【0030】上記ステップS3における処理が終了すると、候補文書格納バッファ5lの内容が参照され、ステップS2で入力されたキーワードをそのテキストデータに含む文書データが存在するか否か、すなわち、候補文書が存在するか否かが調べられる。条件が満たされなかった(候補文書が存在しない)場合には処理はステップS5に、また、条件が満たされた(候補文書が存在する)場合には処理はステップS6にそれぞれ移行する。(ステップS4)。

【0031】ステップS5においては、該当する文書が見つからなかった旨を示すメッセージを表示装置2の表示画面上に表示した後、処理をステップS2に戻してユーザに新たなキーワードを入力させ、上記処理を繰り返す。

【0032】ステップS6においては、候補文書一覧表示部5gが起動され、候補文書一覧表示部5gは候補文書格納バッファ5lに格納されている各文書データのテキストデータ部の内容を参照して、文書ごとにその文書内容表現を表示する。文書内容表現は文字列から構成されており、各文書内容表現は後の処理のために表示装置2の画面上の矩形領域の内部に格納し、この矩形の輪郭を表示する。各文書に対応する文書内容表現を画面上において表示する順序は、候補文書格納バッファ5lに格納されている順序に従う。

【0033】文書内容表現としては、キーワード格納バッファ3jに格納されている文字列を利用して文書データのテキストデータ部からこの文字列を含む箇所が抽出される。抽出される単位は、候補文書数格納バッファ5mに格納されている値によって以下に示す2通りに分類される。

【0034】(1) 候補文書数格納バッファ5mの値が一定値(例えば8)より小さいとき
抽出される単位は段落、つまりテキストデータ部中でCRコードなどで区切られる単位とする。

【0035】(2) 候補文書数格納バッファ5mの値が一定値(例えば8)以上のとき
抽出される単位は文、つまりテキストデータ部中で句点コード(「。」)で区切られる単位とする。

【0036】上記分類(1)、(2)について、キーワードとして「ワークステーション」という語句によって、一つの文書のテキストデータから文書内容表現が抽出される例を原テキストデータと対応づけて図5乃至図7に示す。図5は原テキストデータ10を示し、図6と図7はそれぞれ分類(1)と分類(2)に基づいて抽出されたデータ11,12を示す。

6

【0037】上記したテキストデータ部から文書内容表現を抽出する処理を、候補文書格納バッファ5lに格納されている各文書データに対して実行し、それぞれを文書内容表現として列挙表示する。表示にあたっては、一つの文書データのテキストデータ部に抽出されるテキスト箇所が複数個存在する場合には、その最初に出現する文書内容表現が表示され、また、抽出データにおける該当キーワードが反転表示される。図8に上記分類(2)で抽出された文書内容表現13が各文書ごとに表示されている例を示す。この例においてもキーワードは「ワークステーション」という語句である。

【0038】ステップS6における候補文書一覧の表示の処理が終了すると、文書選択部5hが起動される。文書選択部5hが起動されると、入力装置1のマウス1bを介してユーザによる表示装置2の画面上の位置入力が行なわれる。ここで、ユーザによって指定された位置が、ステップS1で表示されたアイコンと同様の終了コマンドを表すアイコンの内部であれば、一連の検索処理が終了する。(ステップS7、ステップS8)。

【0039】また、ユーザによって指定された位置が、図8に示す文書内容表現13を含む画面上の矩形領域の内部であれば、その矩形が画面上で何番目のものかが調べられ、対応する文書データが候補文書格納バッファ5lから読み出されるとともに文書表示部5iが起動される。文書表示部5iが起動されると、読み出された文書データを構成するテキストデータおよびイメージデータなどが画面上に表示される。文書データの表示処理が終わると、制御はステップS7に戻り、新たな文書データを表示すべく、候補文書一覧に表示されている文書の選択が再度行なわれる。なお、ユーザによって指定された位置が、文書内容表現を含む画面上の矩形領域の外側である場合には、ユーザに正しい位置を指定させるために、ステップS7に戻り、再度位置入力が行なわれる。(ステップS9、ステップS10)。

【0040】なお、上記実施例では候補文書一覧表示部5gにおいて、文書内容表現を抽出時に分類するためのスレシヨールドとして候補文書数格納バッファ5mの値を“8”としたが、これに限ることはなく、表示画面のサイズ、表示文字のサイズなどに対応した最適な値に適宜設定することが可能である。

【0041】また、上記実施例では候補文書一覧表示部5gにおいて、文書内容表現の抽出時に段落単位、あるいは文単位に分類したが、これに限ることはなく、キーワードを含む一文に対して形態素解析を実行して抽出する単位を上記2分類に加えて3分類としたり、あるいは一方の抽出単位と置換しても分類してもよい。

【0042】また、上記実施例では候補文書一覧において、キーワードに該当する語句を反転して表示するようにしたが、キーワードを他の語句と区別できる方法であればいかなる方法であってもよい。例えば、書体、輝

7

度、あるいは色を変えるといった方法でキーワードを表示してもよいのは勿論である。

【0043】また、本発明は上記実施例に限定されるものではなく、本発明の要旨を逸脱しない範囲で種々変形可能であることは勿論である。

【0044】

【発明の効果】以上詳述したように、本発明の文書検索装置によれば、キーワードを用いて検索して得た候補文書の一覧表の要素として、テキスト中の指定されたキーワードを含む箇所を列挙表示することにより、候補として与えられた文書が目的とするものかどうかの判定を瞬時にかつ正確に行なうことができる。その結果、文書データベース中から目的とするものを検索する際に要するユーザの労力を著しく削減することが可能になるなどその実用的効果は多大である。

【0045】また、候補文書の一覧表の要素の表示形態を候補文書の数に応じて変化させることにより、表示画面上での占有面積が小さいのかかわらず、ユーザが目的とする文書を正しく判断するに足る表示となり、操作性の向上を図ることができるという実用上の効果を奏する。

【図面の簡単な説明】

8

【図1】本発明の一実施例の文書検索装置の構成を示すブロック図である。

【図2】文書データ格納装置内における文書データの格納形式を示した図である。

【図3】メモリ装置内部の構成を示した図である。

【図4】処理の流れの概略を示したフローチャートである。

【図5】原テキストデータの例を示す図である。

【図6】分類(1)に基づいて抽出されたデータの例を示す図である。

【図7】分類(2)に基づいて抽出されたデータの例を示す図である。

【図8】文書ごとに文書内容表現が表示されている例を示す図である。

【符号の説明】

1…入力装置 (キーワード入力手段)

3…文書データ格納装置 (文書データ格納手段)

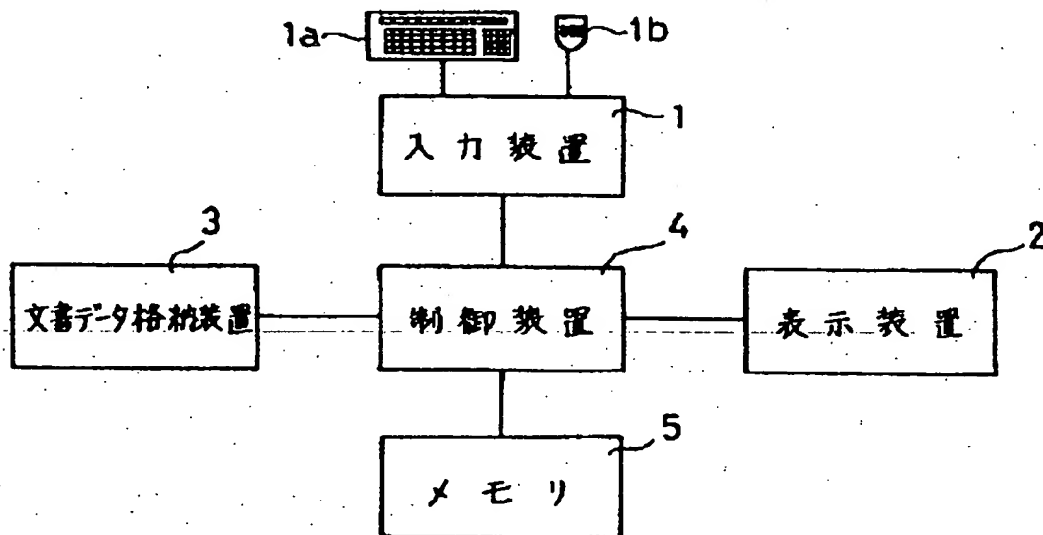
5f…キーワードサーチ部 (キーワードサーチ手段)

5g…候補文書一覧表示部 (文書一覧表示手段)

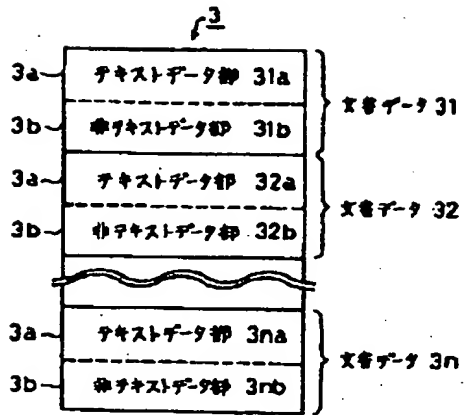
5h…文書選択部 (文書選択手段)

5i…文書表示部 (文書表示手段)

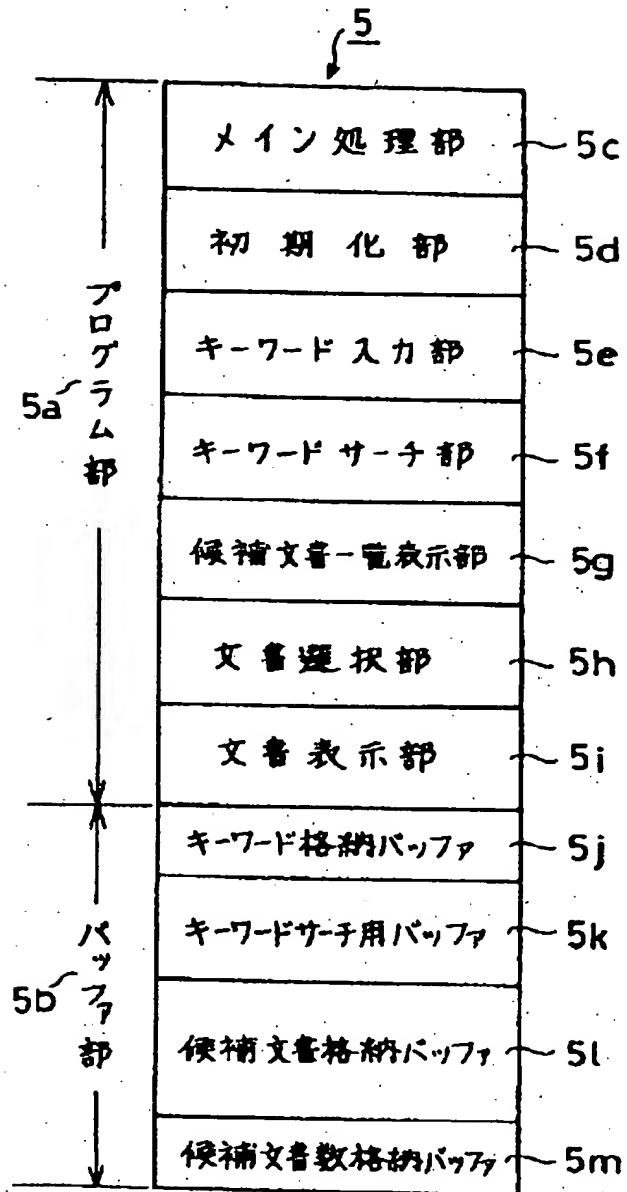
【図1】



【図2】



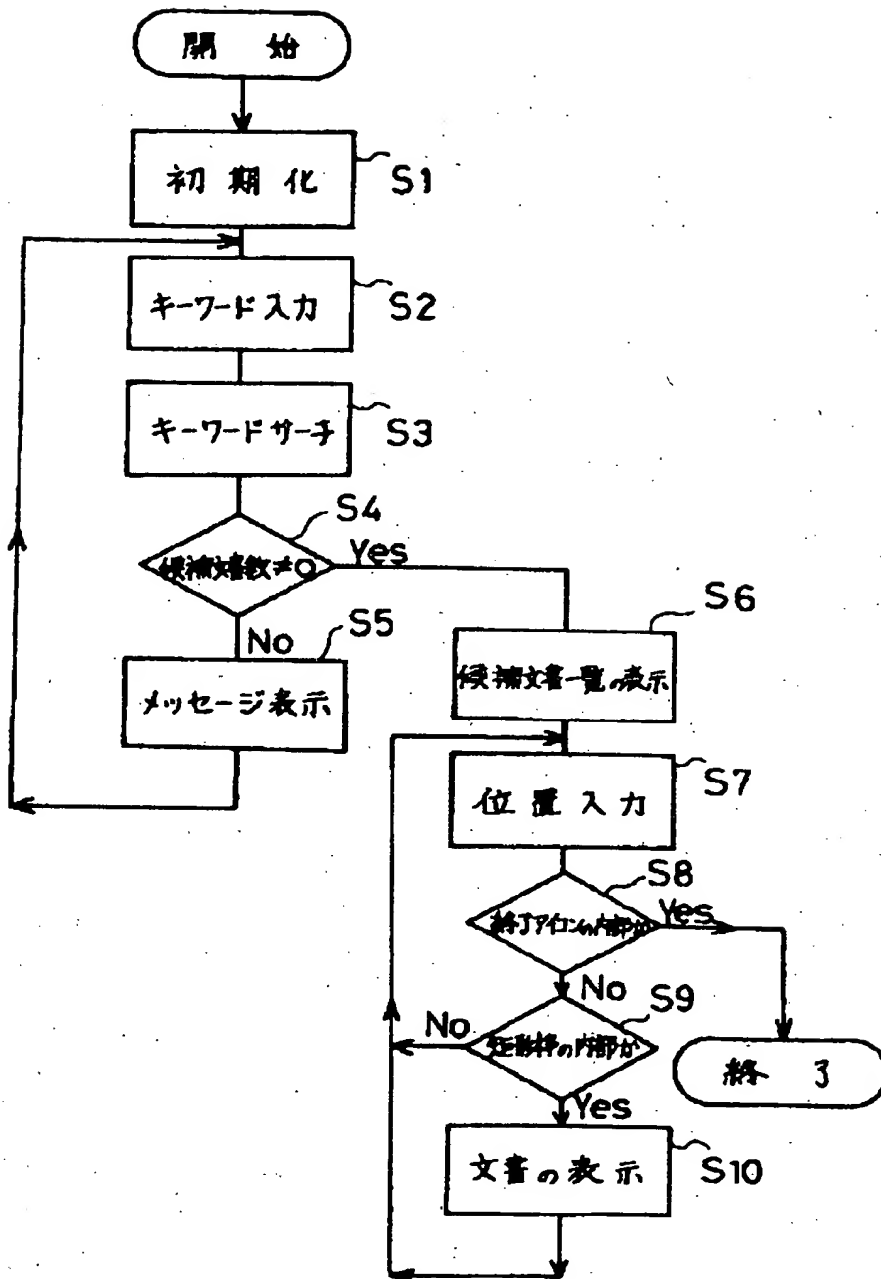
【図3】



【図7】

12
A社は90年2月に、超高速のエンジニアリング・ワークステーションを一般ユーザ向に発売する。

【図4】



【図5】

10

A社は90年2月に、超高速のエンジニアリング・ワークステーションを一般ユーザ向に発売する。このワークステーションはRISCタイプのCPUを2台搭載しており、独自のOSにより高度の分散処理環境を実現している。開発を担当したX氏によると、このワークステーションは特にデータベース処理において、その性能を発揮するということである。

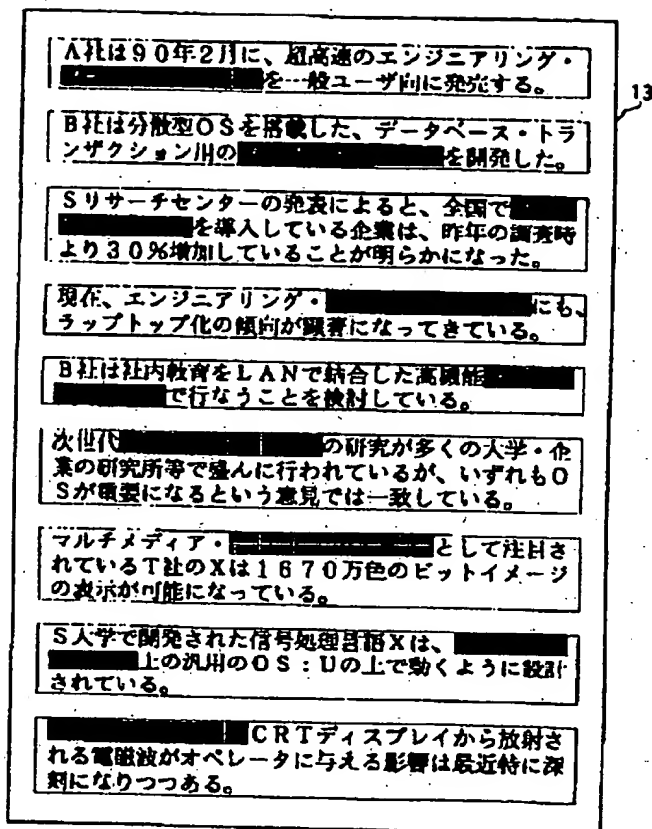
90年4月には、初の有人惑星が種子島宇宙センターから打ち上げられる。まず火星を回る周回軌道を四分の三周回った後、木星の重力圏に突入する。...

【図6】

11

A社は90年2月に、超高速のエンジニアリング・ワークステーションを一般ユーザ向に発売する。このワークステーションはRISCタイプのCPUを2台搭載しており、独自のOSにより高度の分散処理環境を実現している。開発を担当したX氏によると、このワークステーションは特にデータベース処理において、その性能を発揮するということである。

【図8】



**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☒ FADED TEXT OR DRAWING
- ☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.